Digital Signal Processing 27 (2014) 12–22

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

On the design of maximally incoherent sensing matrices for compressed sensing using orthogonal bases and its extension for biorthogonal bases case

Marcio P. Pereira [a], Lisandro Lovisolo [b,*], Eduardo A.B. da Silva [a], Marcello L.R. de Campos [a]

[a] PEE/COPPE/DEL/Poli, Universidade Federal do Rio de Janeiro, Brazil
[b] PROSAICO - PEL/DETEL, Universidade do Estado do Rio de Janeiro, Brazil

ARTICLE INFO

Article history:
Available online 31 January 2014

Keywords:
Compressive sensing
Orthogonal bases
Biorthogonal bases
Signal sampling
Signal compression

ABSTRACT

Compressive Sensing (CS) allows for reconstructing sparse signals within a low acceptable error using less measurements than stipulated by the Nyquist criterion. This CS paradigm rests on the assumption that there is a basis in which the signal is sparse, and one employs random measurements by means of projections on a sensing matrix reconstructing the signal from these measurements through l_1-norm minimization on the sparsity basis. In this work, we propose a method to design sensing matrices with minimum coherence to a given sparsifying orthogonal basis. We provide a mathematical proof of the optimality in terms of coherence minimization for the proposed sensing matrices. This result is extended for biorthogonal bases in order to provide sensing matrices with low coherence, that have advantages when compared to Noiselets in a CS paradigm. Experimental results in an image compression setup show that the proposed sensing matrices provide superior rate-distortion results than Noiselets. These results indicate that the proposed sensing matrices tend to outperform Noiselets when sensing natural images.

1. Introduction

Signal sampling and compression have been around for many years. Recently, a new approach to accomplish these is being investigated: Compressive Sensing (CS) [1–6]. Its objective is to sample signals taking into account their sparsity. For that the signal is projected on known measurement functions or signals. These are designed considering not the signal itself, but instead a basis of the signal space in which the signal is sparse. The signal is then represented by a subset of these measurements. It can be reconstructed from the measurements with an acceptable error by employing optimization algorithms as long as enough measurements are taken. Many previous works have dealt with the reconstruction algorithms and the construction of measurement matrices.

A signal $\mathbf{s} \in \mathbb{R}^N$ is said to be sparse if it can be represented using $K \ll N$ coefficients. For example, if $\mathbf{s}$ has only K non-zero components in an N-dimensional space, it is considered to be sparse and its sparsity K is also referred to as its l_0-norm, the amount of non-zero elements of the vector.

CS exploits the sparsity concept, in the sense that there is a transform Ψ that maps $\mathbf{x}$ into $\mathbf{s} = \Psi\mathbf{x}$ such that $\|\mathbf{s}\|_0 = K$ ($\|\mathbf{s}\|_0$ is the l_0-norm). In order to collect M ($K \leqslant M < N$) samples for representing $\mathbf{x}$ one employs a sensing matrix Φ whose rows are sensing vectors $\boldsymbol{\phi}_m^T$, $m = 1, \ldots, M$. This gives an ensemble of measurements $\mathbf{y}$ such that

$$\mathbf{y} = \Phi\mathbf{x} = \Phi\overline{\Psi}\mathbf{s} \quad \text{or simply} \quad \mathbf{y} = \Theta\mathbf{s}, \tag{1}$$

where $\overline{\Psi}$ is the inverse of Ψ above. Note that Φ and Θ have dimensions $M \times N$.

From the M measurements in $\mathbf{y}$ one then tries to reconstruct $\mathbf{x}$. As it is assumed that $\mathbf{s}$ is sparse one should look for a sparse solution $\hat{\mathbf{s}}$ such that $\mathbf{y} \approx \Theta\hat{\mathbf{s}}$, or, if we consider measurement errors, $\|\mathbf{y} - \Theta\hat{\mathbf{s}}\|_2 < \epsilon$. That is, one should search for

$$\min_{\hat{\mathbf{s}} \in \mathcal{S}} \|\hat{\mathbf{s}}\|_0, \quad \text{subject to} \quad \|\mathbf{y} - \Theta\hat{\mathbf{s}}\|_2 < \epsilon. \tag{2}$$

This is equivalent to looking for

$$\min_{\hat{\mathbf{x}} \in \mathcal{X}} \|\Psi\hat{\mathbf{x}}\|_0, \quad \text{subject to} \quad \|\Phi\hat{\mathbf{x}} - \Theta\hat{\mathbf{s}}\|_2 < \epsilon. \tag{3}$$

Above, $\hat{\mathbf{x}}$ denotes the reconstruction of $\mathbf{x}$. However, the minimization of the l_0-norm is an NP-hard problem. Fortunately, a lot of work has been done to show that this optimization problem can be solved by means of minimizing the l_1-norm of $\hat{s}$, which is a problem computationally more tractable. l_1 minimization [4,6] achieves high probability of reconstructing $\mathbf{x}$ with small errors, i.e., the probability that $\|\hat{\mathbf{x}} - \mathbf{x}\|_2 < \epsilon$ is high, being ϵ a small constant that represents the acceptable measurement error. This probability depends mainly on K, M, N, on the matrices Ψ and Φ, and, obviously, on ϵ.

* Corresponding author.
E-mail addresses: mpp@ieee.org (M.P. Pereira), lisandro@uerj.br (L. Lovisolo), eduardo@smt.ufrj.br (E.A.B. da Silva), campos@smt.ufrj.br (M.L.R. de Campos).

1051-2004/$ – see front matter
http://dx.doi.org/10.1016/j.dsp.2014.01.006

One of the problems often encountered in CS literature is the construction of sensing matrices. The Restricted Isometry Property (RIP) [7] has been the basic foundation for their construction. The RIP is satisfied by some random sensing matrices that have become very popular in CS [5,6]. In [8] criteria that guarantee sensing matrices satisfying an extended RIP are provided. One advantage of random sensing matrices is that they tend to work irrespective of the basis where the vector $\mathbf{x}$ is sparse. For example, if one uses random sensing matrices in an encoder–decoder configuration, only the decoder must be aware of the sparsity basis of $\mathbf{x}$. Notwithstanding the encoder could be aware of the sparsity basis that the decoder uses for recovering the sensed signal. In this case, one could use sensing matrices designed specifically for a given basis in order to obtain improved performance. In [9], a measurement matrix designed from chirp sequences is employed, and properties of it are employed to design a specific reconstruction algorithm. We follow a different approach for constructing such matrices for Compresses Sensing that is applicable for both orthogonal and biorthogonal sparsity bases [10,11].

The coherence between two bases Ψ and Φ of $\mathbb{R}^N$ can be defined as [12]

$$\mu(\Psi,\Phi)=\max_{i,j\in\{0,N-1\}}\frac{|\langle\boldsymbol{\psi}_i,\boldsymbol{\phi}_j\rangle|}{\|\boldsymbol{\psi}_i\|_2\|\boldsymbol{\phi}_j\|_2} \tag{4}$$

where $\boldsymbol{\psi}_i^T$ are the rows of Ψ and $\boldsymbol{\phi}_j^T$ the rows of Φ. The smaller $\mu(\Psi,\Phi)$ is, the more incoherent Ψ and Φ are, i.e. the less similar are the elements of these bases. In [5] it is discussed that in the case of CS considering sparsity in an orthonormal basis Ψ, the sensing matrix Φ should be as incoherent as possible with the basis Ψ.

In this work we present a procedure to construct sensing matrices Φ that have the lowest possible coherence with a given orthogonal sparsity basis Ψ, and extend it to the biorthogonal case for providing sensing matrices with low coherence. We have observed that this approach leads to better rate-distortion performance in case of quantization of the measurements [13,14], when l_1-norm minimization is used to recover the signal.

2. Proposed method for constructing sensing matrices

As mentioned above, a popular way to construct sensing matrices is by using vectors that are, from a statistical viewpoint, incoherent to the basis Ψ [5]. In this work we follow a different approach, by proposing the construction of measurement matrices Φ that are maximally incoherent with Ψ.

2.1. Preliminaries

Theorem 1. *Let $\Psi=\{\boldsymbol{\psi}_1,\boldsymbol{\psi}_2,\ldots,\boldsymbol{\psi}_N\}$ be an orthonormal basis and $\mathbf{h}=\sum_{i=1}^N h_i\boldsymbol{\psi}_i$ be a unit norm vector then $\mathbf{h}$ is maximally incoherent to the basis Ψ if $|h_i|=c$, $\forall i\in\{1,2,\ldots,N\}$.*

Proof. A unit-norm vector $\mathbf{h}$ is maximally incoherent to an orthonormal basis $\Psi=\{\boldsymbol{\psi}_1,\boldsymbol{\psi}_2,\ldots,\boldsymbol{\psi}_N\}$ when it makes the largest angle possible to the closest $\pm\boldsymbol{\psi}_j$, $j\in\{1,\ldots,N\}$.

Since Ψ is a basis, one can write

$$\mathbf{h}=\sum_{j=1}^N h_j\boldsymbol{\psi}_j. \tag{5}$$

Therefore, since $\|\boldsymbol{\psi}_j\|_2=1$ and the basis Ψ is orthogonal, the coherence between the basis Ψ and $\mathbf{h}$ is

$$\mu(\Psi,\mathbf{h})=\max_i\frac{|\langle\boldsymbol{\psi}_i,\sum_{j=1}^N h_j\boldsymbol{\psi}_j\rangle|}{\sum_{j=1}^N|h_j|^2}. \tag{6}$$

Since $\langle\boldsymbol{\psi}_i,\boldsymbol{\psi}_j\rangle=1$ if $i-j=0$ and $\langle\boldsymbol{\psi}_i,\boldsymbol{\psi}_j\rangle=0$ otherwise, we have that

$$\mu(\Psi,\mathbf{h})=\max_i\frac{|h_i|}{\sum_{j=1}^N|h_j|^2}. \tag{7}$$

Therefore, if $\|\mathbf{h}\|_2=1$ then searching for $\min_{\mathbf{h}}\mu(\Psi,\mathbf{h})$ implies solving

$$\mathbf{h}=\operatorname*{argmin}_{\mathbf{h}}\max_i\frac{|h_i|}{\sum_{j=1}^N|h_j|^2},\quad\text{subject to}\quad\sum_{j=1}^N|h_j|^2=1. \tag{8}$$

Since $\sum_{j=1}^N|h_j|^2=1$, this is solved by finding

$$\mathbf{h}=\operatorname*{argmin}_{\{h_1,\ldots,h_N\}}\max|h_i|,\quad\text{subject to}\quad\sum_{j=1}^N|h_j|^2=1. \tag{9}$$

Since the norm of $\mathbf{h}$ is constant, the maximum component h_k will be the lowest when the components are all equal, that is, this can be solved by

$$\langle\boldsymbol{\psi}_i,\mathbf{h}\rangle=|h_i|=c,\quad\forall i\in[1\ldots N], \tag{10}$$

where c is a constant.

The result above implies that vector $\mathbf{h}$ should be equidistant from all the basis elements $\boldsymbol{\psi}_i$. □

For the canonical basis (the axes of $\mathbb{R}^N$), the solution is easy. Any vector in the vertices of the hyper cube (with coordinates $\pm\frac{1}{\sqrt{N}}$) satisfies Eq. (10). This provides 2^N possible solutions. If one has to construct a basis maximally incoherent to Ψ, one can chose N linearly independent elements from the 2^N possible solutions. As it can be noted, the Hadamard basis provides an example of a basis that is most incoherent to the canonical. In addition, any two vectors from this basis have the property of being orthogonal one to another and thus the M vectors in Φ are orthogonal to each other, maximizing information.

2.2. Proposed construction for sensing matrices

Any orthonormal basis can be seen as a rotation of the canonical basis. Therefore, to find a maximally incoherent Φ (an $M\times N$ matrix, $M<N$) to a basis Ψ it suffices to rotate M linearly independent vectors from the hypercube using the rotation that transforms the canonical basis into Ψ. This results in one possible measurement matrix maximally incoherent to Ψ being given by

$$\Phi=H\Psi, \tag{11}$$

where H is such that its elements are $h_{p,q}=\pm\frac{1}{\sqrt{N}}$, $p\in\{1,\ldots,M\}$, $q\in\{1,\ldots,N\}$. One such H can be formed by rows of the Hadamard matrix. Therefore, a simple algorithm expressing this is presented in Algorithm 1.

Algorithm 1 Computing maximally incoherent matrix Φ with M elements w.r.t. an orthogonal basis Ψ.

$H\leftarrow M$ randomly chosen lines from the Hadamard matrix of size N
$\Phi\leftarrow H\Psi$

2.3. Numerical evaluation

Having derived a way to construct maximally incoherent sensing matrices with respect to a given orthonormal basis, in this

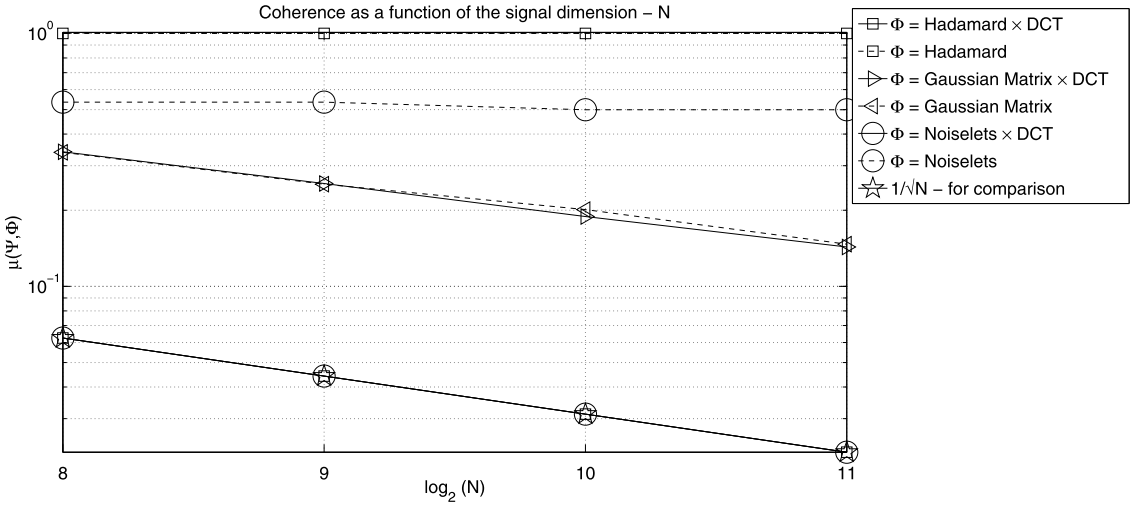


Fig. 1. $\mu(\Psi, \Phi)$ for Ψ being the DCT matrix of dimension N, and different Φ as in the legend box on the right.

subsection we numerically evaluate its performance. We use Algorithm 1 with $M = N$ to construct the maximally incoherent sensing matrices with respect to an orthogonal basis and evaluate the resulting coherence. To evaluate the gain in using the proposed sensing matrices, we compare their coherence with the corresponding orthonormal basis Ψ with the one of commonly employed sensing matrices. For the results here presented Ψ is the DCT basis of the N dimensional space. As Eq. (10) dictates that H in Eq. (11) should have equal entries of constant absolute values, we test two different sets for H, Hadamard basis functions and Noiselets basis functions.

For a fairer comparison, we also present results for a sensing matrix Φ formed of vectors whose elements are drawn from a zero mean unit variance Gaussian distribution that are normalized to unit norm [15,16]. In the remainder of this paper we refer to this set by the name Gaussian. As this set is not deterministic as the Hadamard and the Noiselets ones, we apply this procedure $10 \times 2^{(\log_2(N)-2)}$ times in each case, N being the signal space dimension. Fig. 1 shows the coherence for $N = \{256, 512, 1024, 2048\}$ for both Φ and H in Eq. (11) being the Hadamard, Noiselets and the Gaussian sets. As it can be readily seen from Fig. 1, when $\Phi = H\Psi$ with H satisfying the "constant absolute value coordinates" discussed in Section 2.2, as is the case for Hadamard and Noiselets set, the coherence is minimal, i.e., it equates $\frac{1}{\sqrt{N}}$. These results were obtained for the Hadamard, Noiselets and Gaussian having N linearly independent elements. Note that in Fig. 1 we show results of the Measurement matrix equal to the Gaussian multiplied by the DCT only for completeness. This is so because the Gaussian matrix does not satisfy the restriction of having constant absolute value entries.

As can be seen in Fig. 1, when the Gaussian measurement matrix is employed, it provides a much lower coherence with the DCT basis than the Hadamard and Noiselets measurement matrices (the Hadamard matrix of size N has a DC element vector in the first row as well as the DCT basis). However, when the proposed rotation method (Eq. (11)) is applied then the coherence with the DCT is equal to the theoretical minimum $\frac{1}{\sqrt{N}}$.

Another way to evaluate the coherence of the proposed measurement matrix is to compute the coherence between each vector in $\Phi = \{\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N\}$ and the basis $\Psi = \{\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_N\}$

$$\mu(\boldsymbol{\phi_i}, \Psi) = \max_{\boldsymbol{\psi}_j \in \Psi} \frac{|\langle \boldsymbol{\phi}_i, \boldsymbol{\psi}_j \rangle|}{\|\boldsymbol{\phi}_i\|_2 \|\boldsymbol{\psi}_j\|_2} \tag{12}$$

as $\mu(\Phi, \Psi)$ is the maximum over i of $\mu(\boldsymbol{\phi}_i, \Psi)$. Fig. 2 shows the histogram of this per vector metric for $N = 1024$. The graphs on the right hand side show the coherence of each vector from the DCT matrix with the vectors in Φ for each of the three original Φ, Hadamard in the first row, Gaussian in the middle row and Noiselets in the last row. The vertical line shows the mean of this coherence among all the vectors composing the DCT matrix. The left hand side graphs show the histograms for the modified $\Phi = H\Psi$ – constructed accordingly the proposed procedure. One can readily note the reduction in coherence when the rows of H equate Noiselets and Hadamard basis functions. Also in, these cases one attains the minimal coherence $1/\sqrt{N}$.

3. Sensing matrices for biorthogonal transforms

There is a lot of work on signal processing using biorthogonal transforms. In several applications (e.g. compression) biorthogonal transforms tend to produce better performance than orthogonal ones, as they provide sparser decompositions. This motivated us to investigate the construction of sensing matrices that have low coherence with bases corresponding to biorthogonal transforms. In the previous section we have provided a way of designing sensing matrices that are maximally incoherent to a given orthogonal basis. In what follows, we provide a construction of sensing matrices with low coherence to a given biorthogonal basis. We prove that the proposed sensing functions provide a local minimum on coherence and show through experiments that they indeed possess very low coherence with the biorthogonal basis. We also show via a counter example that the proposed construction does not provide a global incoherence maximum.

In a biorthogonal basis the rows of the sparsifying basis Ψ are not necessarily orthogonal, i.e.

$$\exists i, j \text{ such that } \langle \boldsymbol{\psi}_i, \boldsymbol{\psi}_j \rangle \neq \delta(i - j),\ i \neq j, \tag{13}$$

where $\boldsymbol{\psi}_k^T$ is the k-th row of Ψ and

$$\delta(k) = \begin{cases} 1, & \text{if } k = 0 \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

In this case we define a synthesis basis $\overline{\Psi}$ such that $\overline{\Psi}^T \Psi = I_N$, where I_N is the identity matrix of size N.

Now suppose that the analysis basis is represented by the matrix $\Psi = [\boldsymbol{\psi}_1\ \boldsymbol{\psi}_2\ \ldots\ \boldsymbol{\psi}_N]^T$ and the synthesis basis by the matrix $\overline{\Psi} = [\overline{\boldsymbol{\psi}}_1\ \overline{\boldsymbol{\psi}}_2\ \ldots\ \overline{\boldsymbol{\psi}}_N]^T$. Eq. (5) then becomes the decomposition of vector $\mathbf{h}$ on the biorthogonal basis, that is

$$\mathbf{h} = \sum_{j=1}^{N} h_j \overline{\boldsymbol{\psi}}_j. \tag{15}$$

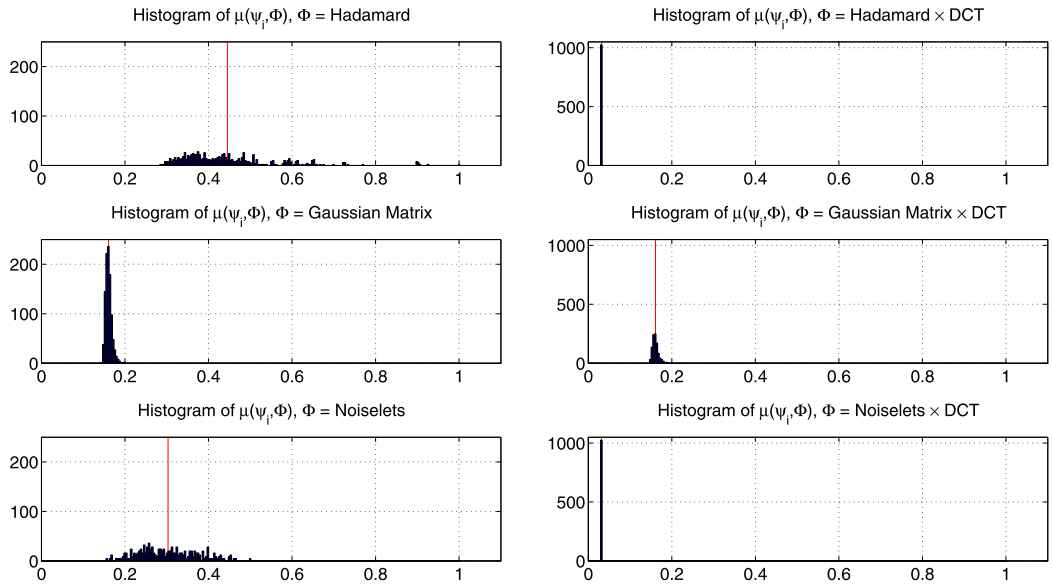


Fig. 2. $\mu(\psi_i, \Phi)$ for Ψ being the DCT matrix of dimension 1024, and different Φ as in the titles.

However, in the biorthogonal case $\overline{\psi}_i$ and $\overline{\psi}_j$ may not be orthogonal, and the derivation presented in Section 2 is no longer valid. In what follows we show a way to build vectors that are locally maximally incoherent with a given biorthogonal basis. For that purpose we will derive a set of new synthesis and analysis bases, $\overline{\psi}'_i$ and ψ'_i respectively. This is explained below.

3.1. Preliminaries

Assuming that $\langle \psi_i, \overline{\psi}_i \rangle = 1$ without any loss of generality, we start by normalizing the synthesis basis so that $\|\overline{\psi}'_i\|_2 = 1$ and $\langle \overline{\psi}'_i, \psi'_i \rangle = 1$, $\forall i$. The net result of this is deriving two matrices Ψ' and $\overline{\Psi}'$ from Ψ and $\overline{\Psi}$, such that $\|\overline{\psi}'_i\|_2 = 1$ and $(\overline{\Psi}')^T \Psi' = I_N$, without changing the directions of the elements.

This can be achieved by normalizing the elements of $\overline{\Psi}$, and scaling the norms of the elements of Ψ accordingly

$$\overline{\psi}'_i = \frac{\overline{\psi}_i}{\|\overline{\psi}_i\|} \quad \text{and} \quad \psi'_i = \|\overline{\psi}_i\| \psi_i. \tag{16}$$

These operations do not change the directions of the vectors defining the basis Ψ nor $\overline{\Psi}$, therefore they do not change their coherence to any set of vectors. Consequently, to build a set of vectors that has low coherence with the biorthogonal basis $\overline{\Psi}$ is equivalent to building a set of vectors that has low coherence with $\overline{\Psi}'$.

From the above, we conjecture that if we could find a vector $\mathbf{h}$ being equidistant from all the vectors of the normalized synthesis basis $\overline{\Psi}'$, then any $\mathbf{h}' \neq \mathbf{h}$ sufficiently close to $\mathbf{h}$ would be closer than $\mathbf{h}$ to one of the basis vectors of $\overline{\Psi}'$, and, therefore, more coherent to them. This implies that vector $\mathbf{h}$ would be, at least locally, maximally incoherent with this basis. Note that this condition is similar to the condition for maximum incoherence for the orthogonal case given by Eq. (8) and Theorem 1. Therefore, we have the theorem below.

Theorem 2. *Let the normalized synthesis basis be* $\{\overline{\Psi}'\}$ $(\|\overline{\psi}'_i\|^2 = 1, \forall i \in \{1, \ldots, N\})$ *and the corresponding analysis basis be* $\{\Psi'\}$. *Let*

$$\mu\left(\mathbf{h}, \{\overline{\psi}'_i\}\right) = \max_i \left|\langle \mathbf{h}, \overline{\psi}'_i \rangle\right| \tag{17}$$

be the similarity/coherence between a vector $\mathbf{h}$ *and the elements of a basis* $\{\overline{\psi}'_i\}$. *For*

$$\mathbf{h} = \sum_{i=1}^{N} \alpha_i \psi'_i \tag{18}$$

the local (inside a small enough N-ball) minimum of $\mu(\mathbf{h}, \{\overline{\psi}'_i\})$ *occurs for* $|\alpha_i| = \alpha$, $\forall i \in \{1, \ldots, N\}$.

The proof of this result is presented in Appendix A.

From the presented derivation, one sees that there are vectors $\mathbf{h}$ locally maximally incoherent with the biorthogonal basis and they are given by

$$\mathbf{h}' = \sum_{j=1}^{N} h_j \psi'_j, \tag{19}$$

where $h_j = \pm\frac{1}{\sqrt{N}}$, $\forall j$. This results derives from the fact that for $\mathbf{h}'$ as in Eq. (19), one has that

$$\left|\langle \mathbf{h}', \overline{\psi}'_j \rangle\right| = \left|\left\langle \sum_{j=1}^{N} h_j \psi'_j, \overline{\psi}'_j \right\rangle\right| = |h_j| = \frac{1}{\sqrt{N}}, \quad \forall j. \tag{20}$$

It should be noted that for the particular case that the basis is orthogonal, this is the condition discussed in Section 2.1.

3.2. Extension of the proposed method for biorthogonal bases

Eq. (19) can be rewritten as

$$\mathbf{h}' = \Psi'^T \mathbf{h} \tag{21}$$

where $\mathbf{h}$ is a vector with entries equal to $\pm\frac{1}{\sqrt{N}}$. Consequently, one can build a sensing matrix with low coherence with Ψ by stacking vectors like $\mathbf{h}^T$ above, yielding

$$\Phi = H\Psi', \tag{22}$$

where H is such that its elements $h_{i,j} = \pm\frac{1}{\sqrt{N}}$, as in the previous section. As in the orthogonal case, the rows of H can be a subset of the Hadamard basis. The algorithm for obtaining Φ locally maximally incoherent to a biorthogonal basis is presented in Algorithm 2.

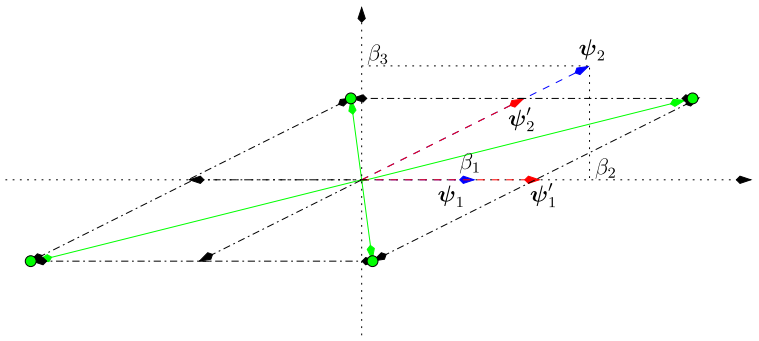


Fig. 3. Example of locally incoherent basis construction for a biorthogonal basis.

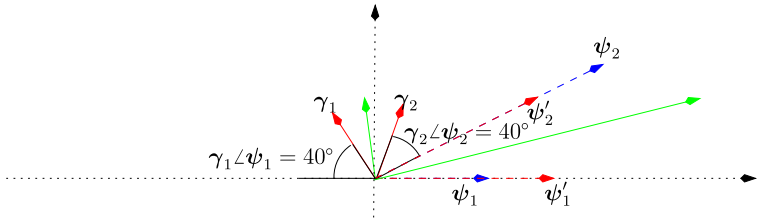


Fig. 4. Example of a set more incoherent with a biorthogonal basis than the one from the construction in Fig. 3.

Algorithm 2 Computing locally maximally incoherent matrix Φ with M elements w.r.t. a biorthogonal basis Ψ.

For $i = 1$ to N

$$\overline{\boldsymbol{\psi}}'_i \leftarrow \frac{\overline{\boldsymbol{\psi}}_i}{\|\overline{\boldsymbol{\psi}}_i\|_2}$$

$$\boldsymbol{\psi}'_i \leftarrow \|\overline{\boldsymbol{\psi}}_i\|_2 \boldsymbol{\psi}_i$$

End

$H \leftarrow M$ randomly chosen lines from the Hadamard matrix of size N

$\Phi \leftarrow H\Psi'$

3.3. Example

We now provide a simple example that shows what the normalizing procedure actually does and which are the vectors that may compose $\Phi = H\Psi'$. For a two dimensional space we have that,

$$\begin{aligned}
&\Phi = H\Psi' \\
&\Phi = \begin{bmatrix} h_{1,1} & h_{1,2} \\ h_{2,1} & h_{2,2} \end{bmatrix} \begin{bmatrix} \psi'_{1,1} & \psi'_{1,2} \\ \psi'_{2,1} & \psi'_{2,2} \end{bmatrix} \\
&\begin{bmatrix} \phi_{1,1} & \phi_{1,2} \\ \phi_{2,1} & \phi_{2,2} \end{bmatrix} = \begin{bmatrix} h_{1,1}\psi'_{1,1} + h_{1,2}\psi'_{2,1} & h_{1,1}\psi'_{1,2} + h_{1,2}\psi'_{2,2} \\ h_{2,1}\psi'_{1,1} + h_{2,2}\psi'_{2,1} & h_{2,1}\psi'_{1,2} + h_{2,2}\psi'_{2,2} \end{bmatrix}.
\end{aligned} \tag{23}$$

That is, each $\boldsymbol{\phi}_i$ is a linear combination of the vectors $\boldsymbol{\psi}'_j$, and the weights $h_{i,j}$ of the $\boldsymbol{\psi}'_j$ have all the same absolute value for H as envisioned.

Now, let the analysis basis Ψ of the 2D space be $\Psi = [\beta_1 \;\; 0; \beta_2\beta_3]$, with $\beta_k \neq 0$, $k = 1, 2, 3$, and as it can be seen the basis vectors are in the rows of Ψ. The inverse to it – the synthesis basis is $\overline{\Psi} = [\frac{1}{\beta_1} \;\; \frac{-\beta_2}{\beta_1\beta_3}; 0 \;\; \frac{1}{\beta_3}]$, such that $\Psi\overline{\Psi}^T = I_2$. In this case Ψ' becomes $\Psi' = [\sqrt{1 + \frac{\beta_2}{\beta_3}} \;\; 0; \;\; \frac{\beta_2}{\beta_3} \;\; 1]$.

Fig. 3 shows the construction of Φ for this basis. In dashed lines we have the vectors of the original basis Ψ. Also, dashed lines show the vectors of Ψ'. In dash-dotted line we show all the possible combinations of ψ' to build the vectors that may compose Φ, which in turn are represented in solid lines. The vectors that can be generated by Algorithm 2 are represented by the circles. All the possible outcomes are shown – for simplicity and without any loss of generality, in this example, the elements of H are considered to have values ± 1 instead of $\frac{\pm 1}{\sqrt{N}}$.

The vectors in Φ obtained using this procedure are inside the convex hull of the set composed by choosing either ψ_i or $-\psi_i$ for $i \in \{1, \ldots, N\}$. It is important to note that, as the proof of Theorem 2 assumes that there is only one vector inside each convex hull, we can only state that the proposed construction generates locally maximally incoherent functions. If we relax this condition, depending on the actual coherence (angle) between the biorthogonal basis elements, one can find vectors that are more incoherent with the basis elements. This is illustrated in Fig. 4 by the vectors $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ – these are obtained directly from $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ to be most incoherent with them, what is possible due to the low dimensionality of the problem. These have an angle of 40° with the basis elements, while the vectors obtained with the proposed procedure have angles of 30° and 60° with the biorthogonal basis elements, and thus a larger coherence since $\cos(30°) > \cos(40°)$.

3.4. Numerical evaluation

We now present some empirical data on the coherence reduction obtained between biorthogonal transforms basis and sensing function vectors constructed accordingly to Algorithm 2. As we have stated in Theorem 2, and exemplified above, Algorithm 2 provides a way to construct locally maximally incoherent sensing functions for biorthogonal bases. In this section we show experimentally that the proposed construction generates sensing vectors with a coherence to a biorthogonal basis that is much lower than the ones of the noiselets and random Gaussian vectors.

We present results using the same framework as in Section 2.3. The basis we use to test the proposed method for constructing low coherent sensing function sets using Algorithm 2 is the CDF 9–7 Biorthogonal DWT [17]. Fig. 5 shows the coherence with the CDF 9–7 for the space dimension $N = \{16, 32, 64, 128, 256, 512, 1024, 2048, 4096\}$ for H being the Hadamard, Noiselets and the Gaussian sets. It also shows the coherence for sets of sensing functions constructed as $\Phi = H\Psi$. As it can be readily seen from Fig. 5 there is a reduction in coherence when the sensing functions are constructed in accordance with Algorithm 2, that is making

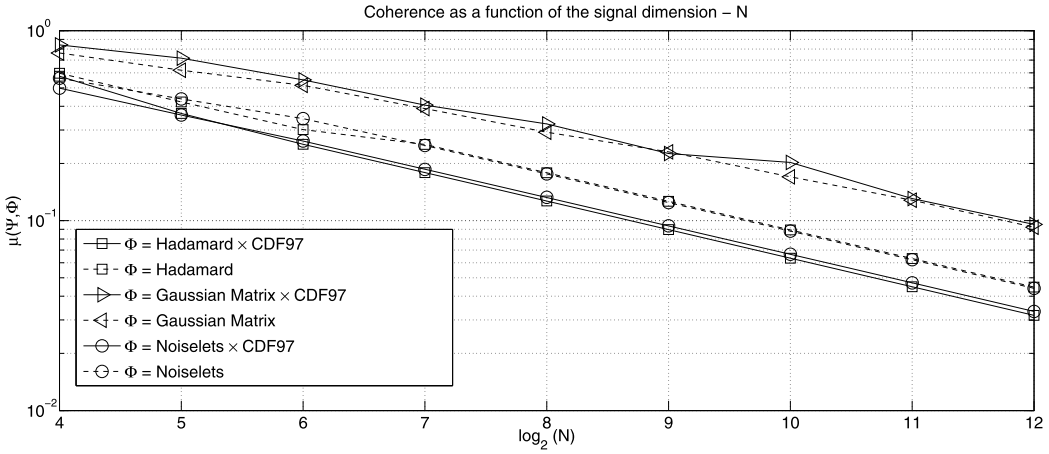


Fig. 5. $\mu(\Psi, \Phi)$ for Ψ being the CDF 9–7 matrix of dimension N, and different Φ as in the legend box on the right.

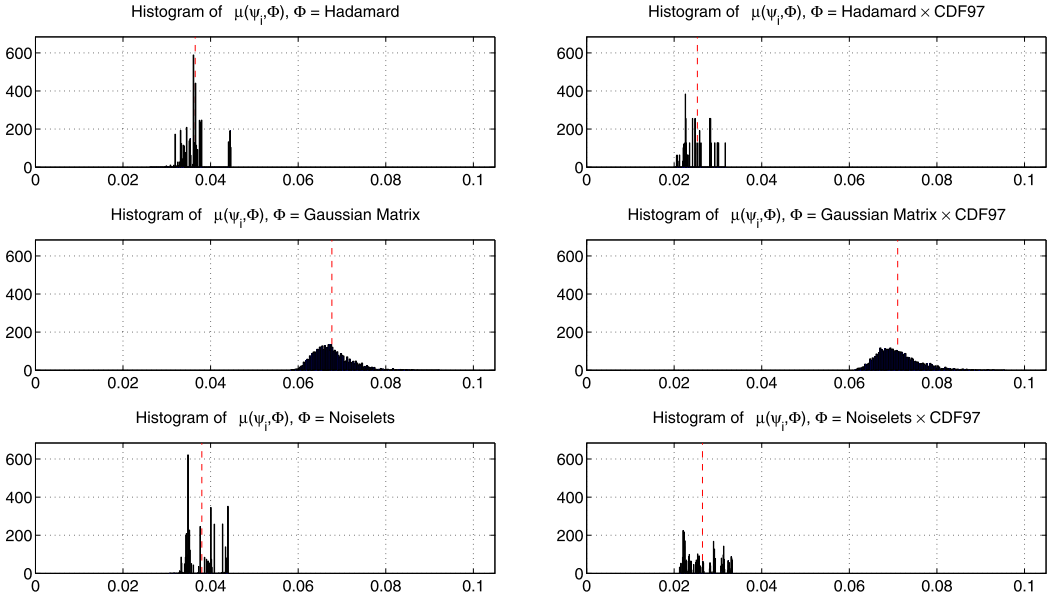


Fig. 6. $\mu(\psi_i, \Phi)$ for Ψ being the CDF 9–7 matrix of dimension 4096, and different Φ as in the titles.

$\Phi = H\Psi$ with H having entries of constant absolute values as claimed in Theorem 2.

Fig. 6 shows the histograms of $\mu(\psi_i, \Phi)$ as defined in Eq. (12) for Ψ being the CDF 9–7 basis of dimension 4096. The graphs on the right hand side show the coherence of each vector from the CDF 9–7 matrix with the vectors in Φ for each of the three original Φ, Hadamard in the first row, Gaussian in the middle row and Noiselets in the last row. The vertical line shows the mean of this coherence among all the vectors composing the CDF 9–7 matrix. The left hand side graphs show the histograms for the modified $\Phi = H\Psi$ – constructed accordingly the proposed procedure. One can readily note the reduction in coherence when H equates the Noiselets and Hadamard bases, thus satisfying Theorem 2.

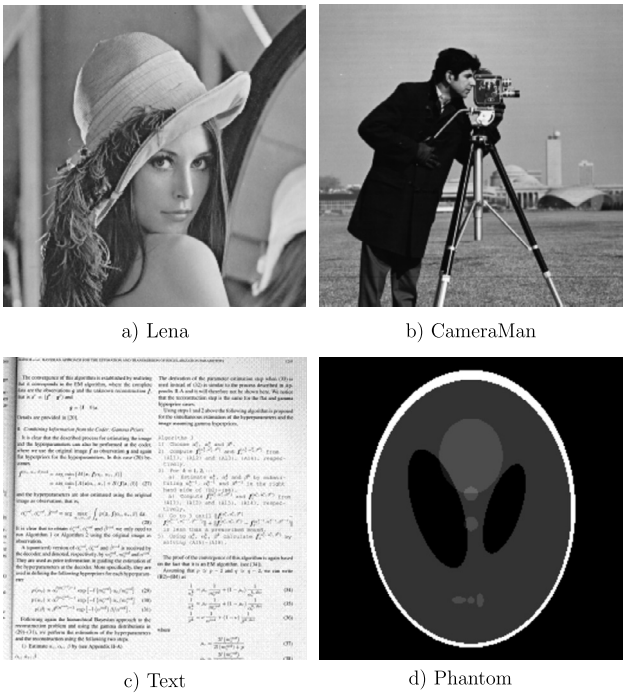
4. Results

For evaluating the proposed approach for constructing measurement matrices we have assumed signal sparsity in both DCT and Biorthogonal DWT (CDF 9–7) bases. As sensing matrices, we have used either the proposed maximally incoherent matrices as defined in Eq. (11) for the orthogonal case and the proposed locally maximally incoherent sensing matrices as in Eq. (22) for the biorthogonal case, and also Noiselets [18,19] for comparison. This results in 4 (four) possible combinations of measurement matrices Φ and sparsity basis Ψ, which are:

- Noiselets with DCT (Orthogonal) (Noiselet@DCT);
- Maximally incoherent sensing matrices with DCT (Orthogonal) (H×DCT@DCT);
- Noiselets with Biorthogonal DWT (CDF 9–7) (Noiselet@CDF97);
- Low incoherent sensing matrices with Biorthogonal DWT (CDF 9–7) with (H×CDF97@CDF97).

The presented selection of sparsifying transform/space (bases) and measurement matrices allows for evaluating the impact of the proposed measurement matrices in comparison to Noiselets, which are very commonly employed in the compressed sensing literature. In this sense, we hope to empirically evaluate the validity of the proposed method for construction of maximally incoherent sensing matrices for orthogonal bases and its extension to construct locally maximally incoherent sensing matrices for biorthogonal bases.

For evaluating the proposed sensing matrices, we have applied the above combinations of sparsity domain × sensing matrices for image compression – a very common Compressive Sensing evaluation approach [4,10,11]. It is important to note that neither we propose this as the application focus for the sensing matrices we



a) Lena b) CameraMan

c) Text d) Phantom

Fig. 7. Original images.

propose nor we intend to have a good image compressor based on CS. This is so because the result presented in this work is theoretical. The image reconstruction application evaluated aims solely to empirically verify its validity. That is, we verify if the reduction in coherence observed numerically in Sections 2.3 and 3.4 using the proposed sensing matrices does provide a gain in this very simple CS evaluation framework. That being said, we simply sense images using different sensing matrices and then reconstruct them from quantized measurements to evaluate the distortion using PSNR (Peak Signal to Noise Ratio) – $10\log_{10}\frac{255^2}{\text{MSE}}$, MSE being Mean Squared Error against the rate spent for encoding the coefficients. This rate simply accounts for the transmission of the quantized measurements as in [14]. For that purpose, images are sensed with varying quantity of measurements. The measurements are then quantized with different levels of distortion, as in [14]. From the quantized measurements the images are reconstructed using l_1-norm minimization.

It is relevant to mention that for convex optimization the `l1eq_pd` routine from `l1-magic` [20] optimization package was used. In it, the primal–dual tolerance (duality gap) was set to 2000, the maximum number of iterations for the primal–dual algorithm was set to 3000, the conjugate gradient tolerance was set to 10^{-8} and the maximum iterations for conjugate gradient was set to 10^4. Although there are several other reconstruction methods in the CS literature [21,22,2,4], we employ l_1-norm minimization since it is not designed considering other signal characteristic besides sparsity, as, for example, is case of the Total Variation (TV) [21,23]. TV is designed considering specific images features. Therefore the reconstruction using TV could hinder gains derived from the proposed sensing matrices since it does not account only for signal reconstruction considering sparsity but also other signal properties and therefore tends to perform better for natural images.

We present results for four images: Lena, CameraMan, Text and Phantom. These images are shown in Fig. 7. The set employed has images with different characteristics and contents as we intend to evaluate the impact of the proposed sensing matrices with images having different characteristics. Lena and CameraMan reproduce natural scenes and due to that, have been used in different papers regarding image processing and compression. Phantom is known to be more sparse and due to that it has been used for evaluation in various compressed sensing papers. Text, being an excerpt of a text page, presents a larger amount of high frequency content.

For each image and for a given quantity of measurements, the measurement–quantization–reconstruction process is applied 10 (ten) times. This provides different realizations of compressive sampling using different measurement vectors from the possible ones (the measurement vectors are draw at random from the set of possible measurements vectors using a uniform probability distribution). More specifically, we have used 20 000, 40 000, 45 000, 50 000, 55 000 and 60 000 measurements. The measurements are then uniformly quantized and the required bit-rate is computed as in [14]. For each image and rate the mean of the attained MSE (Mean Squared Error) for the reconstructed images is used to compute the PSNR (Peak Signal to Noise Ratio).

Figs. 8, 9, 10 and 11 show bit-rate versus peak signal-to-noise ratio (PSNR) results for different combinations of sensing matrices. In these graphs, the symbol ◯ is used to denote that the sensing matrix was constructed according to the algorithms presented while the □ denotes that the sensing matrices are Noiselets. In addition, straight lines denote that the sparsifying basis considered is a CDF97 (biorthogonal wavelet) while dashed lines denote that the sparsifying basis is the DCT (orthogonal). They are used to verify the effectiveness of the proposed method for constructing sensing matrices having low coherence with the sparsifying basis. As previously mentioned, the different bit-rates and PSNR are due to the use of a different number of sensing vectors as well as due to different quantizers applied to the sensed values. Fig. 8 shows results for the image "Lena" 256×256, Fig. 9 for the image "CameraMan" 256×256, Fig. 10 for the image "Text" 256×256, and at last Fig. 11 for the image "Phantom" 256×256.

Some features come out from the graphs:

- In general, assuming a biorthogonal basis as the sparse domain provides larger PSNR in the reconstruction of compressive sensed images than when an orthogonal basis is assumed;
- When the proposed low coherence sensing matrix is employed instead of the ones derived from Noiselets, then for the bases tested the rate-distortion performance tends to improve.

The image Phantom is formed by simple geometric shapes, with highly sparse DCT and Wavelet transforms. We note that, in this case, the difference between the results of Noiselets and the proposed measuring functions is marginal. This is so because, since the level of sparsity is very large, CS has a very good performance, which is not very much dependent on the measuring matrix used. From the results observed one can say that using the low coherence matrices proposed in this work (maximally incoherent in the orthogonal bases case and locally maximally incoherent in the biorthogonal bases case) may be a reasonable and adequate choice for constructing sensing matrices for compressive sensing.

5. Conclusion

In this paper we have proposed sensing matrices for use in Compressive Sensing for a given sparsifying transformation. We have proved that the proposed sensing matrices are maximally incoherent with respect to the sparsifying basis considered in the case the sparsifying basis is orthogonal. In the case the sparsifying basis is biorthogonal we provide locally maximally incoherent sensing matrices. We present proofs of these claims for both, orthogonal and biorthogonal sparsifying bases.

In addition, it should be emphasized that the proposed full rank sensing matrices can be constructed deterministically, that is, directly from the sparsifying basis assumed, without needing to use

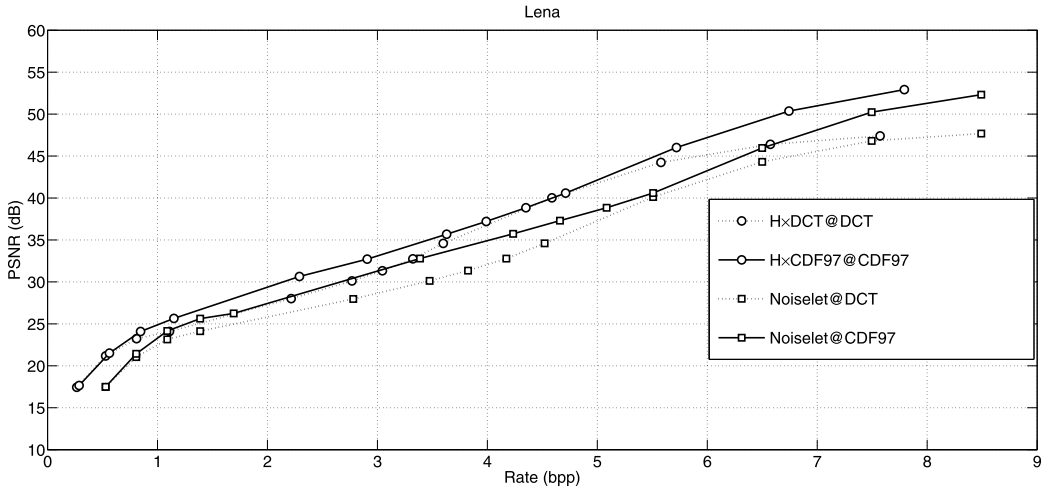


Fig. 8. Bit-rate × PSNR for the natural image "Lena", for different combinations of basis and sensing matrices.

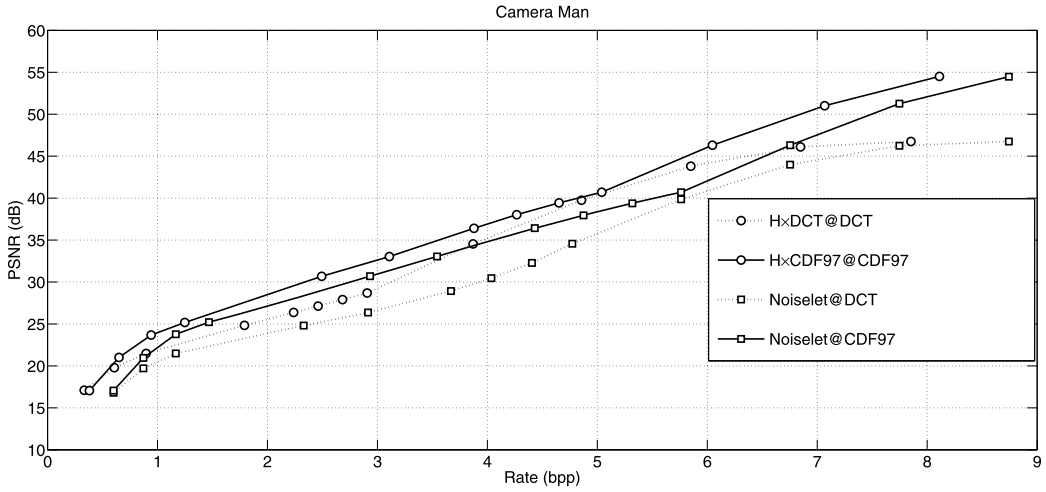


Fig. 9. Bit-rate × PSNR for the natural image "CameraMan", for different combinations of basis and sensing matrices.

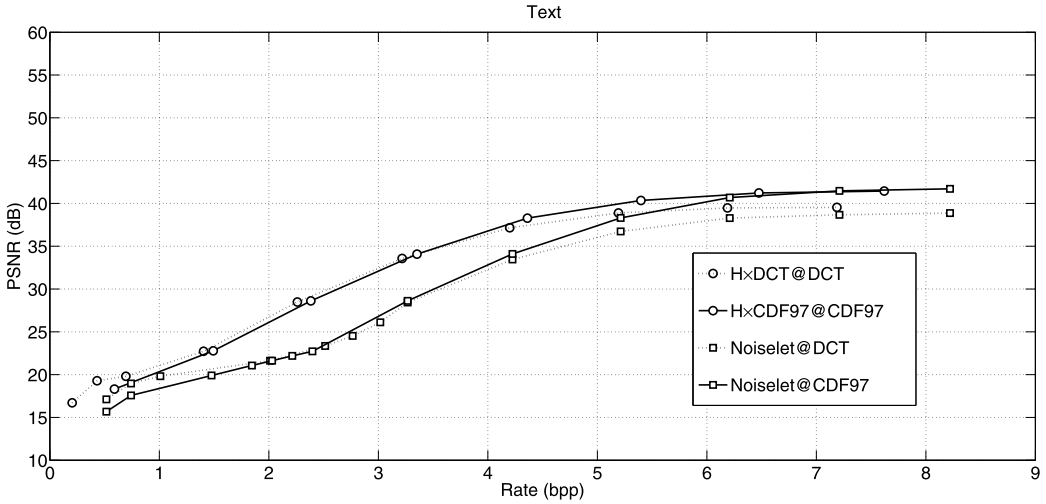


Fig. 10. Bit-rate × PSNR for the image "Text", for different combinations of basis and sensing matrices.

any randomizing algorithm for the construction. An advantage of this approach is that one guarantees incoherence, not in statistical sense as is the case for example when Noiselets are employed, but from the construction itself. Note that, in our approach, the only part that is not deterministic is the choice of the rows of the full rank sensing matrix that will be used as sensing functions. Therefore, one always obtains a measurement matrix being incoherent with the sparsifying basis, maximally for orthogonal bases

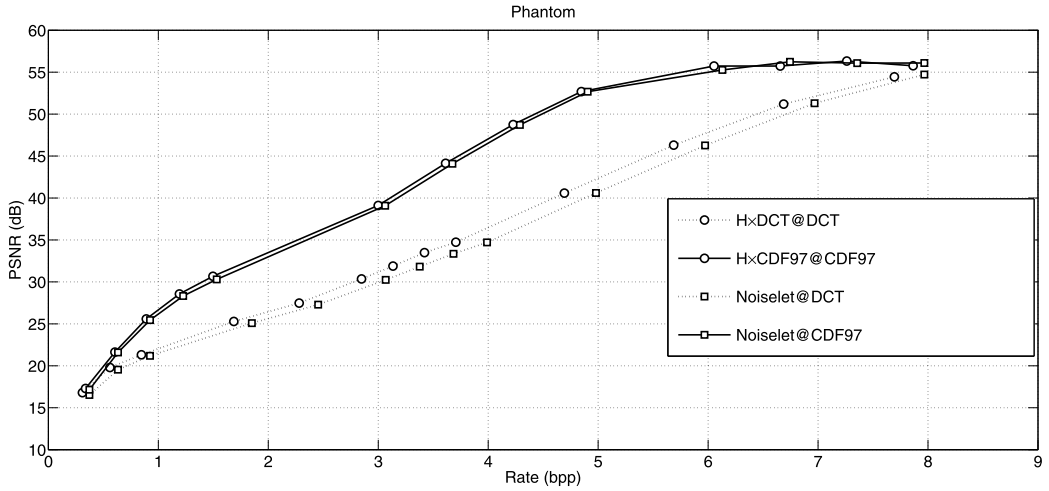


Fig. 11. Bit-rate × PSNR for the synthetic image "Phantom", for different combinations of basis and sensing matrices.

and locally maximally incoherent for biorthogonal bases. Empirical evaluation of the coherence obtained when the proposed sensing matrices are considered is provided that corroborates our claims.

A natural and intuitive test for evaluating the performance of using biorthogonal bases in CS is with images, as for image coding such bases are employed with very good results. Therefore, we have tested our sensing matrices in this framework as a proof of concept – to verify if the proposed sensing matrices were capable of providing gains in actual CS applications and thus corroborate the theoretical results presented. As we have observed from presented results, in image coding applications, the rate-distortion performance of CS using l_1-norm minimization is improved when the proposed sensing matrices are used, both for orthogonal and biorthogonal bases. It is worth noting that, when we use a biorthogonal transform (as the CDF 9–7) as sparsifying basis, locally maximally incoherent sensing matrices lead to better rate-distortion performances than the ones obtained with maximally incoherent sensing matrices for orthogonal transforms as sparsifying basis, probably because images are more sparse in biorthogonal bases than in orthogonal ones.

Therefore, with the presented sensing vectors construction methods of Algorithms 1 and 2, once the sparsifying basis (orthogonal or biorthogonal) is devised for some class of signals then incoherent sensing vectors can be computed and employed for compressed sensing of this class. In addition, the proposed methods to generate sensing vectors for Compressed Sensing do not impact time and space complexity of the encoder–decoder process using Compressed Sensing as the complexity of CS is impacted mainly by the recovery algorithm used. The sensing vectors constructed using the proposed approaches are maximally incoherent in the case of an orthogonal basis and locally maximally incoherent in the case of a biorthogonal basis. This is an indication that these characteristics may provide reconstruction gains as compared to other sets of sensing vectors in CS applications.

Appendix A. Proof of local maximal incoherence of the proposed matrices

Preliminaries. Let $\{\boldsymbol{\psi}'_i\}$ and $\{\overline{\boldsymbol{\psi}}'_i\}$ be two biorthogonal bases of $\mathbb{R}^N$, one has that $\langle \boldsymbol{\psi}'_i, \overline{\boldsymbol{\psi}}'_j \rangle = \delta(i-j)$.

Without any loss of generality one can normalize either, e.g. obtaining the basis $\{\overline{\boldsymbol{\psi}}_i\}$ such that $\|\overline{\boldsymbol{\psi}}_i\|_2 = 1$, $i \in \{1,\ldots,N\}$ and scaling the other accordingly to obtain $\{\boldsymbol{\psi}_i\}$, as in Section 3, and one still has $\langle \boldsymbol{\psi}_i, \overline{\boldsymbol{\psi}}_j \rangle = \delta(i-j)$.

Note that the notation here differs from the one in Section 3; we interchanged $\boldsymbol{\psi}'_i$ for $\boldsymbol{\psi}_i$ in order for the notation to be less cumbersome.

That being said, below, we prove that the bases obtained in Sections 2 and 3 are composed by maximally incoherent vectors.

Lemma. *Let*

$$\mathbf{p} = \sum_{i=1}^{N} \alpha_i \boldsymbol{\psi}_i. \tag{A.1}$$

$\mathbf{p}$ *is equidistant from all* $\overline{\boldsymbol{\psi}}_i$, $i \in \{1,\ldots,N\}$ *if and only if* $\alpha_i = \alpha_j$, $i, j \in \{1,\ldots,N\}$.

Proof. Let

$$d_i = \|\overline{\boldsymbol{\psi}}_i - \mathbf{p}\|_2. \tag{A.2}$$

By definition $d_i^2 = \|\overline{\boldsymbol{\psi}}_i\|_2^2 + \|\mathbf{p}\|_2^2 - 2\overline{\boldsymbol{\psi}}_i^T \mathbf{p}$. However,

$$\overline{\boldsymbol{\psi}}_i^T \mathbf{p} = \mathbf{p}^T \overline{\boldsymbol{\psi}}_i = \sum_{j=1}^{N} \alpha_j \boldsymbol{\psi}_j^T \overline{\boldsymbol{\psi}}_i = \alpha_i \boldsymbol{\psi}_i^T \overline{\boldsymbol{\psi}}_i = \alpha_i. \tag{A.3}$$

Since $\|\overline{\boldsymbol{\psi}}_i\|^2 = 1$, then $d_i^2 = 1 + \|\mathbf{p}\|_2^2 - 2\alpha_i$.
Therefore, $d_i = d_j$ if and only if $\alpha_i = \alpha_j$, $\forall i, j$. □

Corollary A. *For a given value of* $\|\mathbf{p}\|_2$, *the distances from* $\mathbf{p}$ *to the vectors of the basis* $\{\overline{\boldsymbol{\psi}}_i\}$ *are the same if and only if* $\alpha_i = \alpha_j = \alpha > 0$, $\forall i, j \in \{1,\ldots,N\}$.

Proof. It is straightforward since $\|\overline{\boldsymbol{\psi}}_i\|_2 = \|\overline{\boldsymbol{\psi}}_j\|_2 = 1$ and $\overline{\boldsymbol{\psi}}_i^T \mathbf{p} = \alpha_i = \alpha$. □

Corollary B. *Let* $d_i = \|\overline{\boldsymbol{\psi}}_i - \mathbf{p}\|_2$, $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1\ \boldsymbol{\psi}_2\ \cdots\ \boldsymbol{\psi}_N]$ *and* $\mathbf{1}^T = [1\ 1\ \cdots\ 1]$. *If* $\forall i \in \{1,\ldots,N\}$, $\alpha_i = \alpha = \frac{1}{\beta}$, *with*

$$\beta = \mathbf{1}^T \boldsymbol{\Psi}^T \boldsymbol{\Psi} \mathbf{1}, \tag{A.4}$$

then the distances are

$$d_i = \left[\frac{\beta - 1}{\beta}\right]^{1/2}, \quad \forall i \in \{1,\ldots,N\} \tag{A.5}$$

and this is the minimal value attainable by these distances.

Proof. Since the value of α that minimizes d_i is the same that minimizes d_i^2, we minimize the latter.

From the definition of $\mathbf{p}$ in the lemma above, one has that

$$\mathbf{p} = \boldsymbol{\Psi}[\alpha_1\ \alpha_2\ \cdots\ \alpha_N]^T. \tag{A.6}$$

For the case of $\mathbf{p}$ being equidistant from the $\overline{\boldsymbol{\psi}}_i$, from the lemma above one already knows that $\alpha_i = \alpha$, $\forall i \in \{1, \ldots, N\}$, then $\mathbf{p} = \boldsymbol{\Psi}\mathbf{1}\alpha$. Therefore,

$$\|\mathbf{p}\|_2^2 = \mathbf{1}^T\boldsymbol{\Psi}^T\boldsymbol{\Psi}\mathbf{1}\alpha^2. \tag{A.7}$$

The squared distance is given by

$$\begin{aligned} d_i^2 &= \|\overline{\boldsymbol{\psi}}_i\|_2^2 + \|\mathbf{p}_i\|_2^2 - 2\overline{\boldsymbol{\psi}}_i^T\mathbf{p} \\ &= 1 + \|\mathbf{p}_i\|_2^2 - 2\alpha = 1 + \alpha^2\mathbf{1}^T\boldsymbol{\Psi}^T\boldsymbol{\Psi}\mathbf{1} - 2\alpha. \end{aligned} \tag{A.8}$$

Differentiating with respect to α and equating to zero one has that

$$\frac{\partial}{\partial\alpha}\left(d_i^2\right) = 2\alpha\beta - 2 \quad \Rightarrow \quad \alpha = \frac{1}{\beta}. \tag{A.9}$$

Substituting this value in the definition of d_i^2 one obtains that

$$d_i = \left(\frac{\beta - 1}{\beta}\right)^{1/2}. \qquad \square \tag{A.10}$$

Theorem. *Let*

$$\mu\left(\mathbf{p}, \{\overline{\boldsymbol{\psi}}_i\}\right) = \max_i \left|\langle \mathbf{p}, \overline{\boldsymbol{\psi}}_i\rangle\right| \tag{A.11}$$

be the similarity/coherence between $\mathbf{p}$ *and the elements of the basis* $\{\overline{\boldsymbol{\psi}}_i\}$ *(recall that* $\|\overline{\boldsymbol{\psi}}_i\|^2 = 1$, $\forall i\{1, \ldots, N\}$*). For*

$$\mathbf{p} = \sum_{i=1}^{N} \alpha_i \boldsymbol{\psi}_i \tag{A.12}$$

the minimum of $\mu(\mathbf{p}, \{\overline{\boldsymbol{\psi}}_i\})$ *occurs for* $\alpha_i = \alpha$, $\forall i \in \{1, \ldots, N\}$.

Proof. One knows that

$$\langle \mathbf{p}, \overline{\boldsymbol{\psi}}_i\rangle = \left\langle \sum_j \alpha_j \boldsymbol{\psi}_j, \overline{\boldsymbol{\psi}}_i \right\rangle = \sum_j \alpha_j \boldsymbol{\psi}_j^T \overline{\boldsymbol{\psi}}_i = \alpha_i. \tag{A.13}$$

Let $\tilde{\mathbf{p}} = \mathbf{p} + \mathbf{e}$ be a vector such that $\|\tilde{\mathbf{p}}\|_2 = \|\mathbf{p}\|_2$ and $\tilde{\mathbf{p}} \in B(\mathbf{p}, \rho)$, where $B(\mathbf{p}, \rho)$ is the N-ball centered in $\mathbf{p}$ of radius ρ.

From Corollary A we have that for being equidistant to the basis elements one needs $\alpha_i = \alpha$, $\forall i \in \{1, \ldots, N\}$.

One can make ρ sufficiently small so that $\tilde{\mathbf{p}}$ lies inside the cone defined by $\{\boldsymbol{\psi}_i\}$, that is,

$$\tilde{\mathbf{p}} = \sum_i \eta_i \boldsymbol{\psi}_i, \quad \eta_i > 0\ \forall i. \tag{A.14}$$

Then

$$\langle \tilde{\mathbf{p}}, \overline{\boldsymbol{\psi}}_i\rangle = \left\langle \sum_j \alpha \boldsymbol{\psi}_j + \mathbf{e}, \overline{\boldsymbol{\psi}}_i \right\rangle = \sum_j \alpha \boldsymbol{\psi}_j^T \overline{\boldsymbol{\psi}}_i + \mathbf{e}^T \overline{\boldsymbol{\psi}}_i = \alpha + \mathbf{e}^T \overline{\boldsymbol{\psi}}_i. \tag{A.15}$$

From this one derives that

$$\mu\left(\tilde{\mathbf{p}}, \{\overline{\boldsymbol{\psi}}_i\}\right) = \max_i \left|\langle \tilde{\mathbf{p}}, \overline{\boldsymbol{\psi}}_i\rangle\right| = \max_i \left|\alpha + \mathbf{e}^T \overline{\boldsymbol{\psi}}_i\right|. \tag{A.16}$$

Assuming that $\alpha > 0$ (the case $\alpha < 0$ is analogous), for

$$\mu\left(\tilde{\mathbf{p}}, \{\overline{\boldsymbol{\psi}}_i\}\right) < \mu\left(\mathbf{p}, \{\overline{\boldsymbol{\psi}}_i\}\right) \tag{A.17}$$

it is necessary that

$$\mathbf{e}^T \overline{\boldsymbol{\psi}}_i < 0. \tag{A.18}$$

This means that $\mathbf{e}$ should belong to the cone generated by $\{-\boldsymbol{\psi}_i\}$, what leads to

$$\mathbf{e} = \sum_j \gamma_j \boldsymbol{\psi}_j \quad \Rightarrow \quad \overline{\boldsymbol{\psi}}_i^T \mathbf{e} = \sum_j \gamma_j \overline{\boldsymbol{\psi}}_i^T \boldsymbol{\psi}_j = \gamma_i. \tag{A.19}$$

This, in turn, implies that

$$\overline{\boldsymbol{\psi}}_i^T \mathbf{e} < 0 \quad \Leftrightarrow \quad \gamma_i < 0. \tag{A.20}$$

From Eqs. (A.12), (A.14) and (A.19),

$$\mathbf{p} = \sum_i \alpha \boldsymbol{\psi}_i, \qquad \tilde{\mathbf{p}} = \sum_i \eta_i \boldsymbol{\psi}_i, \qquad \mathbf{e} = \sum_i \gamma_i \boldsymbol{\psi}_i \tag{A.21}$$

and one has that

$$\begin{aligned} \tilde{\mathbf{p}} = \mathbf{p} + \mathbf{e} \quad \Rightarrow \quad \sum_i \eta_i \boldsymbol{\psi}_i &= \sum_i \alpha \boldsymbol{\psi}_i + \sum_i \gamma_i \boldsymbol{\psi}_i \\ &= \alpha \sum_i \boldsymbol{\psi}_i + \sum_i \gamma_i \boldsymbol{\psi}_i. \end{aligned} \tag{A.22}$$

By left-multiplying the last expression by $\overline{\boldsymbol{\psi}}_j^T$ and from the fact that $\eta_i > 0$ (Eq. (A.14)) one obtains

$$\eta_j = \alpha + \gamma_j > 0. \tag{A.23}$$

If $\mathbf{e}$ belongs to the cone defined by $\{-\boldsymbol{\psi}_j\}$ then $\gamma_j < 0\ \forall j$, (Eq. (A.20)). Therefore

$$0 < \eta_j < \alpha, \quad \forall j. \tag{A.24}$$

However, since $\|\mathbf{p}\|_2 = \|\tilde{\mathbf{p}}\|_2$, one has that

$$(\alpha\mathbf{1})^T\boldsymbol{\Psi}^T\boldsymbol{\Psi}(\alpha\mathbf{1}) = \mathbf{n}^T\boldsymbol{\Psi}^T\boldsymbol{\Psi}\mathbf{n}, \quad \mathbf{n} = [\eta_1\eta_2\cdots\eta_N]. \tag{A.25}$$

This contradicts Eq. (A.24), because $\boldsymbol{\Psi}\boldsymbol{\Psi}^T$ is positive definite.

Therefore, considering the coherence definition in Eq. (A.11) there is no vector inside the N-ball centered in $\mathbf{p}$ of radius ρ having smaller coherence than $\mathbf{p}$ as defined in Eq. (A.1) with $\alpha_i = \beta\ \forall i \in \{1, \ldots, N\}$, with β as in Eq. (A.4), Corollary B. This concludes the proof. $\square$

References

[1] E. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inf. Theory 52 (2006) 489–509.

[2] D.L. Donoho, Compressed sensing, IEEE Trans. Inf. Theory 52 (2006) 1289–1306.

[3] R. Baraniuk, A lecture on compressive sensing, IEEE Signal Process. Mag. 24 (2007) 118–121.

[4] J. Romberg, Imaging via compressive sampling, IEEE Signal Process. Mag. 25 (2008) 14–20.

[5] E. Candès, R. Wakin, An introduction to compressive sampling, IEEE Signal Process. Mag. 25 (2008) 21–30.

[6] E. Candès, T. Tao, Decoding by linear programming, IEEE Trans. Inf. Theory 51 (2005) 4203–4215.

[7] E. Candès, T. Tao, Near optimal signal recovery from random projections: Universal encoding strategies, IEEE Trans. Inf. Theory 52 (2006) 5406–5425.

[8] R. Calderbank, S. Howard, S. Jafarpour, Construction of a large class of deterministic sensing matrices that satisfy a statistical isometry property, IEEE J. Sel. Top. Signal Process. 4 (2010) 358–374.

[9] L. Applebaum, S.D. Howard, S. Searle, R. Calderbank, Chirp sensing codes: Deterministic compressed sensing measurements for fast recovery, Appl. Comput. Harmon. Anal. 26 (2009) 283–290.

[10] L. Wang, X. Wu, G. Shi, Binned progressive quantization for compressive sensing, IEEE Trans. Image Process. 21 (2012) 2980–2990.

[11] J.E. Fowler, S. Mun, E.W. Tramel, Multiscale block compressed sensing with smoother projected Landweber reconstruction, in: Proceedings of the European Signal Processing Conference, pp. 564–568.

[12] E. Candès, J. Romberg, Sparsity and incoherence in compressive sampling, Inverse Probl. 23 (2007).
[13] J.Z. Sun, V.K. Goyal, Optimal quantization of random measurements in compressed sensing, IEEE Int. Symp. Inf. Theory (2009) 6–10.
[14] A. Schulz, E. da Silva, L. Velho, On the empirical rate-distortion performance of compressive sensing, IEEE Int. Conf. Image Proc. (2009) 3049–3052.
[15] R. Chartrand, V. Staneva, Restricted isometry properties and nonconvex compressive sensing, Inverse Probl. 24 (2008) 035020.
[16] A. Cohen, W. Dahmen, R. DeVore, Compressed sensing and best k-term approximation, J. Am. Math. Soc. 22 (2009) 211–231.
[17] A. Cohen, I. Daubechies, J.-C. Feauveau, Biorthogonal bases of compactly supported wavelets, Commun. Pure Appl. Math. 45 (1992) 485–560.
[18] R. Coifman, F. Geshwind, Y. Meyer, Noiselets, Appl. Comput. Harmon. Anal. 10 (2001) 27–44.
[19] T. Tuma, P. Hurley, et al., On the incoherence of noiselet and Haar bases, in: International Conference on Sampling Theory and Applications, SAMPTA'09, 2009.
[20] E. Candes, J. Romberg, L1-magic: Recovery of Sparse Signals via Convex Programming, 2005.
[21] J. Tang, B.E. Nett, G.-H. Chen, Performance comparison between total variation (TV)-based compressed sensing and statistical iterative reconstruction algorithms, Phys. Med. Biol. 54 (2009) 5781.
[22] E. van den Berg, M.P. Friedlander, Probing the Pareto frontier for basis pursuit solutions, SIAM J. Sci. Comput. 31 (2008) 890–912.
[23] K.T. Block, M. Uecker, J. Frahm, Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint, Magn. Reson. Med. 57 (2007) 1086–1098.

Marcio P. Pereira holds an Electronics Engineer degree from the Federal University of Rio de Janeiro (UFRJ) and a M.Sc. degree in Electrical Engineering from COPPE/UFRJ. His academic works and research interests are focused on digital signal processing, mainly on video and audio compression techniques. He serves as Brazilian Society for Television Engineering (SET) Director since 2012, and works for television networks, designing, building and managing broadcast stations all around Brazil.

Lisandro Lovisolo was born in Neuquen, Argentina, but considers himself Brazilian. He received the Electronics Engineering degree from Universidade Federal do Rio de Janeiro, in 1999, the M.Sc. degree in Electrical Engineering, in 2001, and the D.Sc. degree in Electrical Engineering both from Universidade Federal do Rio de Janeiro (COPPE/UFRJ). Since 2003 he has been with the Department of Electronics and Communications Engineering (the undergraduate dept.), UERJ. He has also been with the Postgraduate in Electronics Program, since 2008. His research interests lie in the fields of digital signal and image processing, communications, and computer science.

Eduardo A.B. da Silva (M'95, SM'05) was born in Rio de Janeiro, Brazil. He received the Electronics Engineering degree from Instituto Militar de Engenharia (IME), Brazil, in 1984, the M.Sc. degree in Electrical Engineering from Universidade Federal do Rio de Janeiro (COPPE/UFRJ), in 1990, and the Ph.D. degree in Electronics from the University of Essex, England, in 1995.

In 1987 and 1988 he was with the Department of Electrical Engineering at Instituto Militar de Engenharia, Rio de Janeiro, Brazil. Since 1989 he has been with the Department of Electronics Engineering (the undergraduate dept.), UFRJ. He has also been with the Department of Electrical Engineering (the graduate studies dept.), COPPE/UFRJ, since 1996. His teaching and research interests lie in the fields of digital signal, image and video processing. In these fields, he has published over 200 peer reviewed papers. He won the British Telecom Postgraduate Publication Prize in 1995, for his paper on aliasing cancellation in subband coding. He is also co-author of the book "Digital Signal Processing – System Analysis and Design", published by Cambridge University Press, in 2002, that has also been translated to the Portuguese and Chinese languages, whose second edition has been published in 2010.

He has served as associate editor of the IEEE Transactions on Circuits and Systems – Part I, in 2002, 2003, 2008 and 2009, of the IEEE Transactions on Circuits and Systems – Part II, in 2006 and 2007, and of Multidimensional, Systems and Signal Processing, Springer, since 2006. He has been a Distinguished Lecturer of the IEEE Circuits and Systems Society, in 2003 and 2004. He was Technical Program Co-Chair of ISCAS2011. He is a senior member of the IEEE, of the Brazilian Telecommunication Society, and also a member of the Brazilian Society of Television Engineering. His research interests lie in the fields of signal and image processing, signal compression, digital TV and pattern recognition, together with its applications to telecommunications and the oil and gas industry.

Marcello L.R. de Campos was born in Niteroi, Brazil, in 1968. He received the Engineering degree (cum laude) from the Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil, in 1990, the M.Sc. degree from COPPE/UFRJ, in 1991, and the Ph.D. degree from the University of Victoria, Victoria, BC, Canada, in 1995, all in electrical engineering. In 1996, he was post-doctoral fellow with the Department of Electronics, School of Engineering, UFRJ, and with the Program of Electrical Engineering, COPPE/UFRJ. From January 1997 until May 1998, he was Associate Professor with the Department of Electrical Engineering (DE/3), Military Institute of Engineering (IME), Rio de Janeiro. He is currently Associate Professor of the Program of Electrical Engineering, COPPE/UFRJ, where he served as Department Vice-Chair and Chair, in the years 2004 and 2005, respectively. From September to December 1998, he was visiting the Laboratory for Telecommunications Technology, Helsinki University of Technology, Espoo, Finland. He served as IEEE Communications Society Regional Director for Latin America, in 2000 and 2001. In 2001, he received a Nokia Visiting Fellowship to visit the Centre for Wireless Communications, University of Oulu, Oulu, Finland. In 2008, he visited Unik, the University Graduate Center of the University of Oslo, Oslo, Norway. Marcello Campos served as Local-Arrangements Co-Chair for GLOBECOM'99, Finance Chair for SPAWC 2008, Plenary Chair for ISCAS 2011, and Technical Co-Chair for the 2013 Brazilian Telecommunications Symposium. He founded and is the current Chair of the IEEE Signal Processing Society Rio de Janeiro Chapter. He has taught over 150 courses on mobile communications in 15 countries. His research interests include adaptive signal processing in general and its application to distributed networks in particular, adaptive beamforming, statistical signal processing, signal processing for communications, mobile and wireless communications, and MIMO systems.